PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:
H04L 29/06, H04M 3/50	A1	WO 97/40611
		(43) International Publication Date: 30 October 1997 (30.10.97)

(21) International Application Number: PCT/US97/03690

(22) International Filing Date: 18 March 1997 (18.03.97)

(30) Priority Data:
08/635,801 22 April 1996 (22.04.96) US

(71) Applicant: AT & T CORP. [US/US]; 32 Avenue of the Americas, New York, NY 10013-2412 (US).

(72) Inventors: BENEDIKT, Michael, Abraham; 1323 W. Willington #1, Chicago, IL 60657 (US). LADD, David, Alan; 4141 Downers Drive, Downers Grove, IL 60515 (US). RAMMING, James, Christopher; Apartment N-103, 350 Sharon Park Drive, Menlo Park, CA 94025 (US). REHOR, Kenneth, G.; 7108 West 35th Street, Berwyn, IL 60402 (US). TUCKEY, Curtis, Duane; 3546 North Reta, Chicago, IL 60657 (US).

(74) Agent: WEINICK, Jeffrey, M.; AT & T Corp., 200 Laurel Avenue, Middletown, NJ 07748 (US).

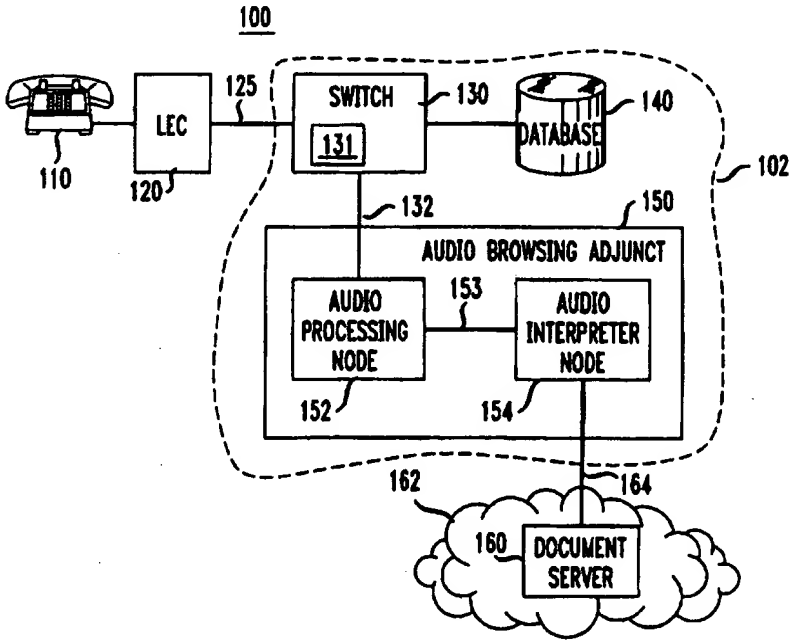
(81) Designated States: CA, IL, JP, KR, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD AND APPARATUS FOR INFORMATION RETRIEVAL USING AUDIO INTERFACE

(57) Abstract

A method and apparatus for retrieving information from a document server (160) using an audio interface device (110). In an advantageous embodiment, a telecommunications network includes an audio browsing node (150) comprising an audio processing node (152) and an audio interpreter node (154). An audio channel is established between the audio interface device and the audio browsing node. A document serving protocol channel (164) is established between the audio browsing node (150) and the document server (160). The document server (160) provides documents to the audio browsing node (150) via the document serving protocol channel (164). The audio browsing node (150) interprets the document into audio data and provides the audio data to the audio interface device (110) via the audio channel. The audio interface device (110) provides audio user input to the audio browsing node (150) via the audio channel. The audio browsing node (150) interprets the audio user input into user data appropriate to be provided to the document server (160) and provides the user data to the document server (160) via the document serving protocol channel (164).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD AND APPARATUS FOR INFORMATION RETRIEVAL USING AUDIO INTERFACE

Field of the Invention

The present invention relates to information retrieval in general. More particularly, the present invention relates to information retrieval over a network utilizing an audio user interface.

Background of the Invention

The amount of information available over communication networks is large and growing at a fast rate. The most popular of such networks is the Internet, which is a network of linked computers around the world. Much of the popularity of the Internet may be attributed to the World Wide Web (WWW) portion of the Internet. The WWW is a portion of the Internet in which information is typically passed between server computers and client computers using the Hypertext Transfer Protocol (HTTP). A server stores information and serves (i.e. sends) the information to a client in response to a request from the client. The clients execute computer software programs, often called browsers, which aid in the requesting and displaying of information. Examples of WWW browsers are Netscape Navigator, available from Netscape Communications, Inc., and the Internet Explorer, available from Microsoft Corp.

Servers, and the information stored therein, are identified through Uniform Resource Locators (URL). URL's are described in detail in Berners-Lee, T., et al., *Uniform Resource Locators*, RFC 1738, Network Working Group, 1994, which is incorporated herein by reference. For example, the URL http://www.hostname.com/document1.html[1], identifies the document

[1] Illustrative URLs are used herein for example purposes only. There is no significance to the use of any particular URL other than for exemplification of the present invention. No reference to actual URLs is intended.

"document1.html" at host server "www.hostname.com". Thus, a request for information from a host server by a client generally includes a URL. The information passed from a server to a client is generally called a document. Such documents are generally defined in terms of a document language, such as Hypertext Markup Language (HTML). Upon request from a client, a server sends an HTML document to the client. HTML documents contain information which is used by the browser to display information to a user at a computer display screen. An HTML document may contain text, logical structure commands, hypertext links, and user input commands. If the user selects (for example by a mouse click) a hypertext link from the display, the browser will request another document from a server.

Currently, WWW browsers are based upon textual and graphical user interfaces. Thus, documents are presented as images on a computer screen. Such images include, for example, text, graphics, hypertext links, and user input dialog boxes. All user interaction with the WWW is through a graphical user interface. Although audio data is capable of being received and played back at a user computer (e.g. a .wav or .au file), such receipt of audio data is secondary to the graphical interface of the WWW. Thus, audio data may be sent as a result of a user request, but there is no means for a user to interact with the WWW using an audio interface.

Summary of the Invention

The present invention provides a method and apparatus for retrieving information from a document server using an audio interface device (e.g. a telephone). An interpreter is provided which receives documents from a document server operating in accordance with a document serving protocol. The interpreter interprets the document into audio data which is provided to the audio user interface. The interpreter also receives audio user input from the audio interface device. The interpreter interprets the audio user input into user data which is appropriate to be sent to the document server in accordance with the document serving protocol and provides the user data to the server . In various embodiments, the interpreter may be located within the audio user interface, within the document server, or disposed in a communication channel between the audio user interface and the document server.

In accordance with one embodiment, a telecommunications network node for carrying out the audio browsing functions of the present invention is included as a node in a telecommunications network, such as a long distance telephone network. An audio channel is established between the audio interface device and the node. A document serving protocol channel is established between the node and the document server. The node receives documents served by the document server in accordance with the document serving protocol and interprets the documents into audio data appropriate for the audio user interface. The node then sends the audio data to the audio interface device via the audio channel. The node also receives audio user input (e.g. DTMF tones or speech) from the audio interface device and interprets the audio user input into user data appropriate for the document server. The node then sends the user data to the document server in accordance with the document serving protocol.

In one embodiment, the document server is a World Wide Web document server which communicates with clients via the hypertext transfer protocol. In accordance with the advantages of the present invention, a user can engage in an audio browsing session with a World Wide Web document server via an audio interface

device. The World Wide Web document server can treat such a browsing session in a conventional manner and does not need to know whether the particular browsing session is being initiated from a client executing a conventional graphical browser or from an audio interface device. The necessary interpreting functions are carried out in the telecommunications network node and these functions are transparent to both a user using the audio interpreting device and the World Wide Web document server operating in accordance with the hypertext transfer protocol.

These and other advantages of the invention will be apparent to those of ordinary skill in the art by reference to the following detailed description and the accompanying drawings.

Brief Description of the Drawings

Fig. 1 shows a diagram of a telecommunications system which is suitable to practice the present invention

Fig. 2 is a block diagram of the components of the audio processing node.

Fig. 3 is a block diagram of the components of the audio interpreter node.

Fig. 4 is a block diagram of a document server.

Fig. 5 is an example audio-HTML document.

Fig. 6 is an example HTML document.

Fig. 7 is a block diagram of an embodiment in which the audio browsing functions are implemented at a user interface device.

Fig. 8 is a block diagram of the components of the user interface device of Fig. 7.

Fig. 9 is a block diagram of an embodiment in which the audio browsing functions are implemented at an audio browsing document server.

Fig. 10 is a block diagram of the components of the audio browsing document server of Fig. 9.

Fig. 11 is a block diagram of an embodiment in which the audio interpreting functions are implemented at an audio interpreter document server.

Fig. 12 is a block diagram of the components of the audio interpreter document server of Fig. 11.

Detailed Description

Fig. 1 shows a diagram of a telecommunications system 100 which is suitable to practice the present invention. An audio interface device, such as telephone 110, is connected to a local exchange carrier (LEC) 120. Audio interface devices other than a telephone may also be used. For example, the audio interface device could be a multimedia computer having telephony capabilities. In accordance with the present invention, a user of telephone 110 places a telephone call to a telephone number associated with information provided by a document server, such as document server 160. In the exemplary embodiment shown in Fig. 1, the document server 160 is part of communication network 162. In an advantageous embodiment, network 162 is the Internet. Telephone numbers associated with information accessible through a document server, such as document server 160, are set up so that they are routed to special telecommunication network nodes, such as audio browsing adjunct 150. In the embodiment shown in Fig. 1, the audio browsing adjunct 150 is a node in telecommunications network 102 which is a long distance telephone network. Thus, the call is routed to the LEC 120, which further routes the call to a long distance carrier switch 130 via trunk 125. Long distance network 102 would generally have other switches similar to switch 130 for routing calls. However, only one switch is shown in Fig. 1 for clarity. It is noted that switch 130 in the telecommunications network 102 is an "intelligent" switch, in that it contains (or is connected to) a processing unit 131 which may be programmed to carry out various functions. Such use of processing units in telecommunications network switches, and the programming thereof, is well known in the art. Upon receipt of the call at switch 130, the call is then routed to the audio browsing adjunct 150. Thus, there is established an

audio channel between telephone 110 and audio browsing adjunct 150. The routing of calls through a telecommunications network is well known in the art and will not be described further herein.

In one embodiment, audio browsing services in accordance with the present invention are provided only to users who are subscribers to an audio browsing service provided by the telecommunication network 102 service provider. In such an embodiment, a database 140 connected to switch 130 contains a list of such subscribers. Switch 130 performs a database 140 lookup to determine if the call originated from a subscriber to the service. One way to accomplish this is to store a list of calling telephone numbers (ANI) in database 140. In a manner which is well known, the LEC 120 provides switch 130 with the ANI of the telephone 110. The switch 130 performs a database 140 lookup to determine if the ANI is included in the list of subscribers to the audio browsing service stored in database 140. If the ANI is present in that list, then the switch 130 routes the call to the audio browsing adjunct 150 in accordance with the present invention. If the ANI does not belong to a subscriber to the audio browsing service, then an appropriate message may be sent to telephone 110.

The audio browsing adjunct 150 contains an audio processing node 152 and an audio interpreter node 154, both of which will be described in further detail below. The audio browsing adjunct 150 provides the audio browsing functionality in accordance with the present invention.

Upon receipt of the call from telephone 110, the audio browsing adjunct 150 establishes a communication channel with the document server 160 associated with the called telephone number via link 164. The association of a telephone number with a document server will be described in further detail below. In a WWW embodiment, link 164 is a socket connection over TCP/IP, the establishment of which is well known in the art. For additional information on TCP/IP, see Comer, Douglas, *Internetworking with TCP/IP: Principles, Protocols, and Architecture*, Englewood Cliffs, NJ, Prentice Hall, 1988, which is incorporated by reference herein. Audio browsing adjunct 150 and the document server 160 communicate with each other

using a document serving protocol. As used herein, a document serving protocol is a communication protocol for the transfer of information between a client and a server. In accordance with such a protocol, a client requests information from a server by sending a request to the server and the server responds to the request by sending a document containing the requested information to the server. Thus, a document serving protocol channel is established between the audio browsing adjunct 150 and the document server 160 via link 164. In an advantageous WWW embodiment, the document serving protocol is the Hypertext Transfer Protocol (HTTP). This protocol is well known in the art of WWW communication and is described in detail in Berners-Lee, T. and Connolly, D., *Hypertext Transfer Protocol (HTTP) Working Draft of the Internet Engineering Task Force*, 1993, which is incorporated herein by reference.

Thus, the audio browsing adjunct 150 communicates with the document server 160 using the HTTP protocol. Thus, as far as the document server 160 is concerned, it behaves as if were communicating with any conventional WWW client executing a conventional graphical browser. Thus, the document server 160 serves documents to the audio browsing adjunct 150 in response to requests it receives over link 164. A document, as used herein, is a collection of information. The document may be a static document in that the document is pre-defined at the server 160 and all requests for that document result in the same information being served. Alternatively, the document could be a dynamic document, whereby the information which is served in response to a request is dynamically generated at the time the request is made. Typically, dynamic documents are generated by scripts, which are programs executed by the server 160 in response to a request for information. For example, a URL may be associated with a script. When the server 160 receives a request including that URL, the server 160 will execute the script to generate a dynamic document, and will serve the dynamically generated document to the client which requested the information. The use of scripts to dynamically generate documents is well known in the art.

The documents served by server 160 include text, logical structure commands, hypertext links, and user input commands. One characteristic of these documents is that the physical structure of the information contained in the document (i.e., the physical layout view of the information when displayed at a client executing a conventional graphics browser), is not defined. Instead, a document contains logical structure commands, which are interpreted at a browser to define a physical layout. For example, such logical structure commands include emphasis commands, new paragraph commands, etc. The syntactic structure of such commands may conform to the conventions of a more general purpose document structuring language, such as Standard Generalized Markup Language (SGML), which is described in Goldfarb, Charles, *The SGML Handbook*, Clarendon Press, 1990, which is incorporated by reference herein. In the WWW embodiment of the present invention, these documents are Hypertext Markup Language (HTML) documents. HTML is a well known language based on SGML which is used to define documents which are served by WWW servers. HTML is described in detail in Berners-Lee, T. and Connolly, D., *Hypertext Markup Language (HTML), Working Draft of the Internet Engineering Task Force,* 1993, which is incorporated herein by reference.

When an HTML document is received by a client executing a conventional browser, the browser interprets the HTML document into an image and displays the image upon a computer display screen. However, in accordance with the principles of the present invention, upon receipt of a document from document server 160, the audio browsing adjunct 150 converts the document into audio data. The details of such conversion will be discussed in further detail below. The audio data is then sent to telephone 110 via switch 130 and LEC 120. Thus, in this manner, the user of telephone 110 can access information from document server 160 via an audio interface.

In addition, the user can send audio user input from the telephone 110 back to the audio browsing adjunct 150. This audio user input may be, for example, speech signals or DTMF tones. The audio browsing adjunct 150 converts the audio user input into user data or instructions which are appropriate for transmitting to the

document server 160 via link 164 in accordance with the HTTP protocol. The user data or instructions are then sent to the document server 160 via the document serving protocol channel. Thus, user interaction with the document server is via an audio user interface.

In this manner, a user can engage in a browsing session with a WWW document server via an audio interface. The document server can treat such a browsing session in a conventional manner and does not need to know whether a particular browsing session is being initiated from a client executing a conventional graphical browser or from an audio interface such as a telephone. The audio browsing adjunct 150 within the network 102 interprets the documents being served by document server 160 into audio data appropriate to be sent to telephone 110. In addition, the audio browsing adjunct 150 interprets audio user input received at telephone 110 into user data appropriate to be received by the document server 160.

A more detailed description of an advantageous embodiment will now be given in conjunction with an example browsing session. Assume a user at telephone 110 dials the number (123) 456-7890[2] which has been set up to be associated with information accessible through document server 160 and therefore routed to audio browsing adjunct 150. The call gets routed to LEC 120, at which point LEC 120 recognizes the telephone number as one which is to be routed to long distance network 102, and more particularly to switch 130. Upon receipt of the call, switch 130 in turn routes the call to the audio browsing adjunct 150 via link 132. Thus, there is established an audio channel between telephone 110 and audio browsing adjunct 150.

Further details of the audio processing node 152 are shown in Fig. 2. The audio processing node 152 comprises a telephone network interface module 210, a DTMF decoder/generator 212, a speech recognition module 214, a text to speech module 216, and an audio play/record module 218, each of which is connected to an

[2] Telephone numbers are used herein for example purposes only. There is no significance to the use of any particular telephone number other than for exemplification of the present invention. No reference to actual telephone numbers is intended.

audio bus 220 and a control/data bus 222, as shown in Fig. 2. Further, the audio processing node 152 contains a central processing unit 224, memory unit 228, and a packet network interface 230, each of which is connected to the control/data bus 222. The overall functioning of the audio processing node 152 is controlled by the central processing unit 224. Central processing unit 224 operates under control of executed computer program instructions 232 which are stored in memory unit 228. Memory unit 228 may be any type of machine readable storage device. For example, memory unit 228 may be a random access memory (RAM), a read only memory (ROM), a programmable read only memory (PROM), an erasable programmable read only memory (EPROM), an electronically erasable programmable read only memory (EEPROM), a magnetic storage media (i.e. a magnetic disk), or an optical storage media (i.e. a CD-ROM). Further, the audio processing node 152 may contain various combinations of machine readable storage devices, which are accessible by the central processing unit 224, and which are capable of storing a combination of computer program instructions 232 and data 234.

The telephone network interface module 210 handles the low level interaction between the audio processing node 152 and telephone network switch 130. In one embodiment, module 210 consists of one or more analog tip/ring loop start telephone line terminations. Through module 210, central processing unit 224 is able to control link 132 via control data bus 222. Control functions include on-hook/off-hook, ring detection, and far-end on-hook detection. In an alternate embodiment, module 210 includes one or more channelized digital interfaces, such as T1/DS1, E1, or PRI. Signaling can be in-band or out-of-band. The DTMF decoder/generator 212 handles the conversion of DTMF tones into digital data and the generation of DTMF tones from digital data. The speech recognition module 214 performs speech recognition of speech signals originating at user telephone 110 and received over the audio bus 220. Such speech signals are processed and converted into digital data by the speech recognition module 214. The text to speech module 216 converts text of documents received from document server 160 into audio speech signals to be transmitted to a user at telephone 110. The audio play/record module 218 is used to play audio data

received from document server 160 at telephone 110 and to record audio data such as a user's voice. It is noted that each module 210, 212, 214, 216, 218 are shown as separate functional modules in Fig. 2. The functionality of each of modules 212, 214, 216, and 218 may be implemented in hardware, software, or a combination of hardware and software, using well known signal processing techniques. The functionality of module 210 may be implemented in hardware or a combination of hardware and software, using well known signal processing techniques. The functioning of each of these modules will be described in further detail below in conjunction with the example. The packet network interface 230 is used for communication between the audio processing node 152 and the audio interpreter node 154.

The audio browsing adjunct 150 also contains an audio interpreter node 154 which is connected to the audio processing node 152. The audio interpreter node 154 is shown in further detail in Fig. 3. Audio interpreter node 154 contains a central processing unit 302, a memory 304, and two packet network interfaces 306 and 308 connected by a control/data bus 310. The overall functioning of the audio interpreter node 154 is controlled by the central processing unit 302. Central processing unit 302 operates under control of executed computer program instructions 312 which are stored in memory unit 304.

Memory unit 304 may be any type of machine readable storage device. For example, memory unit 304 may be a random access memory (RAM), a read only memory (ROM), a programmable read only memory (PROM), an erasable programmable read only memory (EPROM), an electronically erasable programmable read only memory (EEPROM), a magnetic storage media (i.e. a magnetic disk), or an optical storage media (i.e. a CD-ROM). Further, the audio interpreter node 154 may contain various combinations of machine readable storage devices, which are accessible by the central processing unit 302, and which are capable of storing a combination of computer program instructions 312 and data 314.

The control of an apparatus, such as the audio processing node 152 and the audio interpreter node 154, using a central processing unit executing software instructions is well known in the art and will not be described in further detail herein.

Returning now to the example, the call placed from telephone 110 to telephone number (123) 456-7890, has been routed to the audio browsing adjunct 150, and in particular to the audio processing node 152. The central processing unit 224 detects the ringing line through the telephone network interface module 210. Upon detection of the call, the central processing unit performs a lookup to determine the URL which is associated with the dialed number (DN). The dialed telephone number (DN), is provided to switch 130 from the local exchange carrier 120 in a manner which is well known in the art, and in turn, the DN is provided to the audio browsing adjunct 150 from switch 130. A list of URL's which are associated with DN's is stored as data 234 in memory 228. Assume in the present example the DN (123) 456-7890 is associated with URL http://www.att.com/~phone/greeting.

In an alternate embodiment, the list of URL's associated with various DN's is stored in a network database, such as database 140, instead of locally at the audio browsing adjunct 150. In such an embodiment, the central processing unit 224 of the audio processing node 152 sends a signal to network switch 130 to request a lookup to database 140. The switch would request the URL from database 140 and return the resulting URL to the audio processing node 152. It is noted that the communication between the audio processing node 152, switch 130 and database 140, may be via an out of band signaling system, such as SS7, which is well known in the art. An advantage to this configuration is that a plurality of audio browsing adjuncts may be present in the network 102, and each may share a single database 140. In this manner, only one database 140 needs to be updated with URLs and associated DNs.

After receiving the URL associated with the DN, the central processing unit 224 of the audio processing node 152 sends a message (including the URL) to the audio interpreter node 154 instructing the audio interpreter node 154 to initiate an audio interpreting/browsing session. Such a message is passed from the central processing unit 224 to the packet network interface 230 via the control/data bus 222.

The message is sent from packet network interface 230 of the audio processing node 152 to the packet network interface 306 of the audio interpreting node 154 via connection 153. In an advantageous embodiment, the audio processing node 152 and the audio interpreter node 154 are collocated and thus form an integrated audio browsing adjunct 150. In alternate embodiments, the audio processing node 152 and the audio interpreter node 154 may be geographically separated. Several such alternate embodiments are described below. The connection 153 may be a packet data network connection (e.g., TCP/IP connection over Ethernet) which is well known in the art.

Returning now to the example, the audio interpreter node 154 receives a message via packet network interface 306 that it is to initiate a new audio interpreting/browsing session. The central processing unit 302 is capable of controlling multiple audio interpreting/browsing sessions for multiple users simultaneously. Such multiprocess execution by a processor is well known, and generally entails the instantiation of a software process for controlling each of the session. Upon the initiation of an audio interpreting/ browsing session, the audio interpreting node 154 sends an HTTP request for URL http://www.att.com/~phone/greeting to the document server 160 over connection 164. In this example, it is assumed that the document server 160 is associated with the host name www.att.com.

Document server 160 is shown in further detail in Fig. 4. Document server 160 is a computer containing a central processing unit 402 connected to a memory 404. The functions of the document server 160 are controlled by the central processing unit 402 executing computer program instructions 416 stored in memory 404. In operation, the document server 160 receives requests for documents from the audio interpreter node 154 via connection 164 and packet network interface 440. The central processing unit 402 interprets the requests and retrieves the requested information from memory 404. Such requests may be for HTML documents 408, audio-HTML documents 410, audio files 412, or graphics files 414. HTML documents 408 are well known and contain conventional HTML instructions for use

in conventional WWW graphical browsers. An audio-HTML document is similar to an HTML document but has additional instructions which are particularly directed to interpretation by the audio interpreter node 154 in accordance with the present invention. Such instructions which are particular to the audio browsing aspects of the present invention will be identified herein as audio-HTML instructions. The details of audio-HTML documents and audio-HTML instructions will be described in further detail below. Audio files 412 are files which contain audio information. Graphics files 414 are files which contain graphical information. In a manner which is well known in the art, a URL identifies a particular document on a particular document server. Memory 404 may also contain scripts 418 for dynamically generating HTML documents and audio-HTML documents. Thus, returning to the present example, an HTTP request for URL http://www.att.com/~phone/greeting is received by the document server 160 from the audio interpreter node 154 via connection 164.

The document server interprets this URL and retrieves an audio-HTML page from memory 404 under central processing unit 402 control. The central processing unit 402 then sends this audio-HTML document to the audio interpreter node 154 via packet network interface 440 and link 164.

The audio-HTML document 500 which is sent in response to the request for URL http://www.att.com/~phone/greeting, and which is received by the audio interpreter node 154, is shown in Fig. 5. The audio interpreter node 154 begins interpreting the document 500 as follows. In one embodiment the <HEAD> section, lines 502-506, of the document 500, including the title of the page, is not converted into voice, and is ignored by the audio interpreter node 154. In alternate embodiments, the <TITLE> section may be interpreted using text to speech as described below.

The text "Hello!" at line 508 in the <BODY> section of the document 500 is sent from the audio interpreter node 154 to the audio processing node 152 via packet network interface 306 and link 153. Along with the text "Hello!", the audio interpreter node 154 sends instructions to the audio processing node 152 that the text is to be processed by the text to speech module 216. The audio processing node 152

receives the text and instructions via the packet network interface 230, and the text is supplied to the text to speech module 216 via control/data bus 222. The text to speech module 216 generates the audio signal to play *"Hello"*[3] and sends the signal to the telephone network interface module 210 via audio bus 220. The telephone network interface module 210 then sends the audio signal to telephone 110. It is noted that text to speech conversion is well known and conventional text to speech techniques may be used by the text to speech module 214. For example, the punctuation "!" in the text may be interpreted as increased volume when the text is converted to speech.

Line 510 of document 500 is a form instruction, and the audio interpreter node 154 does not send anything to the audio processing node 152 in connection with this instruction. The audio interpreter node 154 interprets line 510 to indicate that it will be expecting a future response from the user, and that this response is to be given as an argument to the script identified by http://machine:8888/hastings-bin/getscript.sh. Line 512 is an audio-HTML instruction. The audio interpreter node 154 interprets line 512 by sending an http request to server 160 for the audio file identified by www-spr.ih.att.com/~hastings/annc/greeting.mu8, which resides in memory 404 in storage area 412. The document server 160 retrieves the audio file from memory 404 and sends it to the audio interpreter node 154 via link 164. Upon receipt of the audio file, the audio interpreter node 154 sends the file, along with instructions indicating that the file is to be played by the audio play/record module 218, to the audio processing node 152. Upon receipt of the file and instructions, the audio processing node 152 routes the audio file to the audio play/record module 218. The audio play/record module 218 generates an audio signal which is sent to the telephone network interface module 210 via audio bus 220. The telephone network interface module 210 then sends the audio signal to the telephone 110. As a result, the user at telephone 110 hears the contents of the audio file www-spr.ih.att.com/~hastings/annc/greeting.mu8 at the speaker of telephone 110.

Lines 514-516 are audio-HTML instructions. The audio interpreter node 154 does not send line 514 to the audio processing node 152. Line 514 indicates that a

[3] *Italic* type is used herein to indicate text which is played as audio speech.

response from the user is to be sent to the document server 160 associated with the variable name "collectvar". This instruction marks the beginning of a prompt-and-collect sequence in which the user will be prompted for, and supply, information. This instruction is followed by a prompt instruction 516 and a set of choice instructions 518-522. The audio interpreter node 154 processes line 516 in a manner similar to that of line 512, and as a result, the user at telephone 110 hears the audio from the file identified by http://www-spr.ih.att.com/~hastings/annc/choices.mu8. The audio will ask the user to make a selection based upon some criteria, and the audio interpreter node 154 will wait for a response from the user at telephone 110. Also, as a result of processing line 516, the central processing unit 302 sends a message to the audio processing node 152 instructing the telephone network interface module 210 to be prepared to receive audio input.

The user responds with audio user input from telephone 110. The audio user input may be in the form of DTMF tones generated by the user pressing a key on the keypad of telephone 110. For example, if the user presses "2" on telephone 110 keypad, the DTMF tone associated with "2" is received by the audio processing node 152 via the telephone network interface module 210. The audio signal is recognized as a DTMF tone by the central processing unit 224, and instructions are passed to telephone network interface module 210 to send the signal to the DTMF decoder/generator 212 via the audio bus 220. The central processing unit 224 instructs the DTMF decoder/generator 212 to convert the DTMF tone into digital data and to pass the digital data to the packet network interface 230 for transmission to the audio interpreter node 154. Upon receipt of the signal, the audio interpreter node 154 recognizes that the user has responded with choice 2, which corresponds with the value "Jim" as indicated by line 520 of the audio-HTML document 500. Thus, the audio interpreter node 154 sends the value "Jim" associated with the variable "collectvar" to the script http://machine:8888/hastings-bin/getscript.sh identified in line 510 of document 500. If the user responds with input which is not listed as a choice, in this example, a response other than 1-3, or if the user does not respond within a certain time period, then the audio interpreter node 154 instructs the text to

speech module 216 to generate a speech signal "*choice not understood, try again*", and that signal is provided to the user at telephone 110.

Alternatively, audio user input may be in the form of a voice signal. Instead of the user pressing the number 2 on telephone 110 keypad, the user will speak the word "two" into the telephone 110 microphone. The voice signal is received by the audio processing node 152 via the telephone network interface module 210. The audio signal is recognized as a voice signal by the central processing unit 224, and instructions are passed to telephone network interface module 210 to send the signal to the speech recognition module 214 via the audio bus 220. The central processing unit 224 instructs the speech recognition module 214 to convert the voice signal into digital data and to pass the data to the packet network interface 230 for transmission to the audio interpreter node 154. Upon receipt, the audio interpreter node 154 processes the data as described above in conjunction with the DTMF audio user input. It is noted that the speech recognition module 214 operates in accordance with conventional speech recognition techniques which are well known in the art.

Hypertext links often appear in HTML documents. When displayed on the screen of a computer executing a conventional graphical browser, a hypertext link will be graphically identified (e.g. underlined). If a user graphically selects a link, for example by clicking on the link with a mouse, then the browser generates a request for the document indicated by the link and sends the request to the document server. Consider the HTML document 600 shown in Fig. 6. Lines 604 and 605 specify a conventional HTML description of hypertext links. If this page were being processed by a conventional graphical browser, the display would look like:

This page gives you a choice of links to follow to other World Wide Web pages. Please click on one of the links below.

click here for information on cars
click here for information on trucks

The user would then select one of the links using a graphical pointing device such as a mouse. If the user selects the link click here for information on cars then the browser would generate a request for the document identified by the URL http://www.abc.com/cars.html. If the user selects the link click here for information on trucks then the browser would generate a request for the document identified by the URL http://www.abc.com/trucks.html.

The processing of HTML hypertext links in accordance with the present invention will now be described with reference to Fig. 6. Assume that the document server 160 has served the HTML document 600 shown in Fig. 6 to the audio interpreter node 154. Lines 602 and 603 will be converted to audio signals by the text to speech module 216 and provided to the user telephone 110 as described above. Thus, the user will hear the audio, *This page gives you a choice of links to follow to other World Wide Web pages. Please click on one of the links below.* When line 604 is reached, the audio interpreter node 154 will recognize line 604 as being a hypertext link. The audio interpreter node 154 sends an instruction to the audio processing node 152, instructing the DTMF decoder/generator 212 to generate a tone to the telephone 110. Alternatively, the tone could be generated by the audio interpreter node 154 sending an instruction to the audio processing node 152, instructing the audio play/record module 218 to play an audio file containing tone audio. The particular tone is one which is used to signify the beginning of a hypertext link to the user. The audio interpreter node 154 then supplies the text of the hypertext link, "click here for information on cars", to the audio processing node 154 with an instruction indicating that the text is to be processed by the text to speech module 216. As a result, the speech audio signal "*click here for information on cars*", is provided to the telephone 110. The audio interpreter node 154 then sends an instruction to the audio processing node 152, instructing the DTMF decoder/generator 212 to generate a tone to the telephone 110. This particular tone is one which is used to signify the end of a hypertext link to the user. The tones used to signify the beginning and end of hypertext links may be the same or different tones. The ending tone is followed by a

pause. As an alternative to using tones, the beginning and end of a hypertext link may be identified by speech audio signals such as "*begin link [hypertext] end link*".

If the user wishes to follow the link, then the user supplies user audio input during the pause. For example, suppose the user wanted to follow the link "click here for information on cars". The user would enter audio input during the pause following the generated speech audio signal for the link. The audio input may be, for example, a DTMF tone generated by pressing a key on the telephone 110 keypad. The DTMF tone is received by the audio processing node 152 and processed by the DTMF decoder/generator 212. Data representing the DTMF tone is provided to the audio interpreter node 154 via the control/data bus 222, packet network interface 230, and link 153. Upon receipt of the signal, the audio interpreter node 154 recognizes that the signal has been received during the pause following the selected link, and the audio interpreter node 154 generates a request for the WWW document identified by the URL http://www.abc.com/cars.html, which is associated with the selected link. Alternatively, audio user input for selecting a hypertext link may be in the form of a speech signal.

Another type of link is a hypertext anchor link. An anchor link allows a user to jump to a particular location within a single HTML document. In conventional graphical browsers, when a user selects an anchor link, the browser displays the portion of the document indicated by the link. In accordance with the audio browsing techniques of the present invention, if a user selects an anchor link, the audio interpreter node 154 will begin interpreting the document at the point specified by the link. For example, line 620 of document 600 contains a hypertext anchor to the portion of the document at line 625. This hypertext link is identified to the user in a manner similar to that of the hypertext links which identify new HTML documents, as described above. The hypertext anchor links may be distinguished by, for example, a different audio tone or a generated speech signal identifying the link as an anchor link. If the user selects the anchor link at line 620, then the audio interpreter node 154 will skip down to the text at line 625 and will begin interpreting the HTML document 600 at that point.

The advantageous embodiment described above in conjunction with Fig. 1 is configured such that the audio browsing adjunct 150, including the audio processing node 152 and the audio interpreter node 154, is embodied in a telecommunications network node located within a long distance telecommunications network 102. This configuration provides the advantage that the audio browsing functions in accordance with the present invention can be provided to telephone network subscribers by the telephone network 102 service provider. In such a configuration, there is no additional hardware required at the user premises or at the document server. All audio browsing functions are provided by components within the telephone network 102. However, alternate configurations are possible and such alternate configurations could be readily implemented by one skilled in the art in view of the present disclosure.

One such alternate configuration is shown in Fig. 7, in which the functions of the audio browsing adjunct are shown implemented at a user interface device 700. In such an embodiment, the functions of the audio processing node 152, along with the functions of the audio interpreter node 154, are integrated within the single user interface device 700. The user interface device 700 communicates with the document server 160 through a communication link 702. Link 702 is similar to link 164 which was described above in connection with Fig. 1. Thus, link 702 may be a socket connection over TCP/IP, the establishment of which is well known in the art. User interface device 700 is shown in further detail in Fig. 8. User interface device 700 comprises a keypad/keyboard 802 and a microphone 804 for accepting user input, and a speaker 806 for providing audio output to the user. The user interface device 700 also comprises a keypad keyboard interface module 816 connected to a control/data bus 824. The user interface device 700 also comprises a codec 810, a speech recognition module 818, a text to speech module 820, and an audio play/record module 822, each of which is connected to an audio bus 808 and the control/data bus 824 as shown in Fig. 8. The codec 810 contains an analog to digital converter 812 and a digital to analog converter 814, both of which are controlled by a central processing unit 826 via the control/data bus 824. The analog to digital converter 812 converts analog audio user input from microphone 804 into digital audio signals and provides

the digital audio signals to the audio bus 808. The digital to analog converter 814 converts digital audio signals from the audio bus 808 to analog audio signals to be sent to the speaker 806. The keypad/keyboard interface module 816 receives input from the keypad/keyboard 802 and provides the input to the control data bus 824. The speech recognition module 818, the text to speech module 820, and the audio play/record module 822, perform the same functions, and are similarly configured, as modules 214, 216, and 218, respectively, which were described above in conjunction with Fig. 2. In addition, the user interface device 700 contains a packet network interface 834 for connecting to a packet network, such as the Internet, via link 702. Further, the user interface device 700 contains central processing unit 826 and a memory unit 828, both of which are connected to the control/data bus 824. The overall functioning of the user interface device 700 is controlled by the central processing unit 826. Central processing unit 826 operates under control of executed computer program instructions 830 which are stored in memory unit 828. Memory unit 828 also contains data 832.

The user interface device 700 implements the functions of the audio processing node 152 and the audio interpreter node 154, which were described above in conjunction with the embodiment of Fig. 1. These functions are implemented by the central processing unit 826 executing computer program instructions 830. Thus, the computer program instructions 830 would include program instructions which are the same as, or similar to: 1) computer program instructions 232 implementing the functions of the audio processing node 152; and 2) computer program instructions 312 implementing the functions of the audio interpreter node 154. The functioning of the audio processing node 152 and the audio interpreter node 154 were described in detail above, and will not be described in further detail here. Central processing unit 836 is capable of executing multiple processes at the same time, and in this way implements the functions of the audio processing node 152 and the audio interpreter node 154. This multiprocess functioning is illustrated in Fig. 8 where the central processing unit 826 is shown executing audio interpreting/browsing process 836 and audio processing process 838.

In operation, a user of user interface device 700 would request a URL using keypad/keyboard 802 or microphone 804. If the keypad/keyboard 802 is used to request a URL, the keypad/keyboard interface module 816 would provide the requested URL to the central processing unit 826 via the control/data bus 824. If the microphone 804 is used to request a URL, the user's voice is received by microphone 804, digitized by analog to digital converter 812, and passed to the speech recognition module 818 via the audio bus 808. The speech recognition module 818 would then provide the requested URL to the central processing unit 826 via the control/data bus 824.

Upon receipt of the URL, the central processing unit 826 initiates an audio browsing/interpreting session by instantiating an audio interpreting/browsing process 836. The audio interpreting/browsing process 836 sends an HTTP request to the document server 160 via the packet network interface 834 in a manner similar to that described above in conjunction with the embodiment of Fig. 1. Upon receipt of the document from document server 160, the audio interpreting/browsing process 836 interprets the document in accordance with the audio browsing techniques of the present invention. The audio resulting from the interpretation of the document is provided to the user via the speaker 806 under control of the audio processing process 838. Similarly, a user of the user interface device 700 can provide audio user input to the user interface device via the microphone 804.

Since the audio interpreting/browsing process 836 and the audio processing process 838 are co-resident in the user interface device 700, all communications between the two processes takes place through the central processing unit 826 via inter-process communication, and all communication between the processes 836, 838 and other elements of the user interface device 700 takes place via the control/data bus 824.

Figs. 7 and 8 show the user interface device 700 communicating directly with the document server 160 in the packet network 162. Alternatively, the user interface device 700 could be configured to communicated with the document server 160 via a standard telephone connection. In such a configuration, the packet network interface

834 would be replaced with a telephone interface circuit, which would be controlled by central processing unit 826 via control/data bus 824. User interface device 700 would then initiate a telephone call to the document server via the telephone network. The document server 160 would terminate the call from the user interface device 700 using hardware similar to the telephone network interface module 210 (Fig. 2). Alternatively, the call could be terminated within the telephone network, with the termination point providing a packet network connection to the document server 160.

In an alternate configuration shown in Fig. 9, the functions of the audio browsing adjunct 150 (including the functions of the audio processing node 152 and the audio interpreter node 154) and the document server 160 are implemented within an audio browsing document server 900. As illustrated in Fig. 9, calls are routed from a telephone 110, through LEC 120, switch 130, and another LEC 902, to the audio browsing document server 900. Thus, in this particular embodiment, the audio browsing document server 900 could be reached from a conventional telephone 110 via a telephone network. In addition, the audio browsing document server 900 is also connected to the Internet via a link 904. The audio browsing document server 900 is shown in further detail in Fig. 10. The audio browsing document server 900 comprises a telephone network interface module 1010, a DTMF decoder/generator 1012, a speech recognition module 1014, a text to speech module 1016, and an audio play/record module 1018, each of which is connected to an audio bus 1002 and a control/data bus 1004, as shown in Fig. 10. Each of these modules 1010, 1012, 1014, 1016, and 1018 perform the same functions, and are similarly configured, as modules 210, 212, 214, 216, and 218, respectively, which were described above in conjunction with Fig. 2. In addition, the audio browsing document server 900 contains a packet network interface 1044 for connecting to a packet network, such as the Internet, via link 904. The packet network interface 1044 is similar to the packet network interface 230 described above in conjunction with Fig. 2. Further, the audio browsing document server 900 contains a central processing unit 1020 and a memory unit 1030, both of which are connected to the control/data bus 1004. The overall functioning of the audio browsing document server 900 is controlled by the central processing unit

1020. Central processing unit 1020 operates under control of executed computer program instructions 1032 which are stored in memory unit 1030. Memory unit 1030 also contains data 1034, HTML documents 1036, audio-HTML documents 1038, audio files 1040, and graphics files 1042.

The audio browsing document server 900 implements the functions of the audio processing node 152, the audio interpreter node 154, and the document server 160, which were described above in conjunction with the embodiment of Fig. 1. These functions are implemented by the central processing unit 1020 executing computer program instructions 1032. Thus, the computer program instructions 1032 would include program instructions which are the same as, or similar to: 1) computer program instructions 232 implementing the functions of the audio processing node 152; 2) computer program instructions 312 implementing the functions of the audio interpreter node 154; and 3) computer program instructions 416 implementing the functions of the document server 160. The functioning of the audio processing node 152, the audio interpreter node 154, and the document server 160 were described in detail, and will not be described in further detail here. Central processing unit 1020 is capable of executing multiple processes at the same time, and in this way implements the functions of the audio processing node 152, the audio interpreter node 154, and the document server 160. This multiprocess functioning is illustrated in Fig. 10 where the central processing unit 1020 is shown executing audio interpreting/browsing process 1022, document serving process 1024, and audio processing process 1026.

In operation, a call placed by telephone 110 to a telephone number associated with information accessible through the audio browsing document server 900, is routed to the audio browsing document server 900 via LEC 120, switch 130, and LEC 902. It is noted that a plurality of telephone numbers may be associated with various information accessible through the audio browsing document server 900, and each such telephone number would be routed to the audio browsing document server 900. The ringing line is detected through the telephone network interface module 1010 under control of the audio processing process 1026. Upon detection of the call, the central processing unit 1020 performs a lookup to determine the URL which is

associated with the dialed number (DN). The DN is provided to the audio browsing document server 900 from the LEC 902 in a manner which is well known in the art. A list of DN's with associated URL's is stored as data 1034 in memory 1030. Upon receipt of the URL associated with the DN, the central processing unit 1020 initiates an audio browsing/interpreting session by instantiating an audio interpreting/browsing process 1022. The audio interpreting/browsing process 1022 sends an HTTP request to the document serving process 1024 which is co-executing on the central processing unit 1020. The document serving process 1024 performs the document server functions as described above in conjunction with document server 160 in the embodiment shown in Fig. 1. These document server functions are supported by the HTML documents 1036, audio-HTML documents 1038, audio files 1040, and graphics files 1042 stored in memory 1030. Thus, the central processing unit 1020 retrieves the document associated with the URL from memory 1030. The audio interpreting/browsing process 1022 then interprets the document in accordance with the audio browsing techniques of the present invention. The audio resulting from the interpretation of the document is provided to the user under control of the audio processing process 1026. Similarly, a user of telephone 110 can provide audio user input to the audio browsing document server 900 in a manner similar to that described above in conjunction with the embodiment of Fig. 1.

Since the audio interpreting/browsing process 1022, the document serving process 1024, and the audio processing process 1026, are co-resident in the audio browsing document server 900, all communications between the processes 1022, 1024, 1026, takes place through the central processing unit 1020 via inter-process communication, and all communication between the processes 1022, 1024, 1026, and other elements of the audio browsing document server 900 takes place via the control/data bus 1004. One advantage of this embodiment is efficiency, in that HTML documents and other data does not need to traverse a potentially unreliable wide-area network in order to be processed (e.g. interpreted).

In the embodiment shown in Fig. 1, the audio processing node 152 and the audio interpreter node 154 were collocated. However, the functions of the audio

processing node 152 and the audio interpreter node 154 may be geographically separated as shown in Fig. 11. In such an embodiment, the audio processing node 152 is contained within the telecommunications network 102 and an audio interpreter document server 1100 is contained within the packet network 162. The functioning of the audio processing node 152 is as described above in conjunction with the embodiment of Fig. 1. The audio interpreter document server 1100, which implements the functions of a document server, such as document server 160, and the functions of the audio interpreter node 154, is shown in further detail in Fig. 12. The audio interpreter document server 1100 contains a packet network interface 1202 connected to link 153 and to a control/data bus 1204. The audio interpreter document server 1100 contains a central processing unit 1206 and a memory unit 1212, both of which are connected to the control/data bus 1204. The overall functioning of the audio interpreter document server 1100 is controlled by the central processing unit 1206. Central processing unit 1206 operates under control of executed computer program instructions 1214 which are stored in memory unit 1212. Memory unit 1212 also contains data 1216, HTML documents 1218, audio-HTML documents 1220, audio files 1222, and graphics files 1224.

The audio interpreter document server 1100 implements the functions of the audio interpreter node 154 and the document server 160, which were described above in conjunction with the embodiment of Fig. 1. These functions are implemented by the central processing unit 1206 executing computer program instructions 1214. Thus, the computer program instructions 1214 would include program instructions which are the same as, or similar to: 1) computer program instructions 312 implementing the functions of the audio interpreter node 154; and 2) computer program instructions 416 implementing the functions of the document server 160. The functioning of the audio interpreter node 154 and the document server 160 were described in detail above, and will not be described in further detail here. Central processing unit 1206 is capable of executing multiple processes at the same time, and in this way implements the functions of the audio interpreter node 154 and the document server 160. This multiprocess functioning is illustrated in Fig. 12 where the

central processing unit 1206 is shown executing audio interpreting/browsing process 1208 and document serving process 1210.

In operation, the audio processing node 152 communicates with the audio interpreter document server 1100 over link 153 in a manner similar to that described above in conjunction with Fig. 1. However, unlike Fig. 1, in which the audio interpreter node 154 communicated with the document server via link 164, the audio interpreter browsing process 1208 communicates with the document serving process 1210 through the central processing unit 1206 via inter-process communication.

Thus, as described above, the audio browsing aspects of the present invention may be implemented in various ways, such that the audio processing functions, the audio interpreting/browsing functions, and the document serving functions, may be integrated or separate, depending on the particular configuration. One skilled in the art would recognize that there are other possible configurations for providing the audio browsing functions of the present invention.

As can be seen from the above description, the present invention may be used in conjunction with standard HTML documents, which are generally intended to be used with conventional graphics browsers, or with audio-HTML documents which are created specifically for use in accordance with the audio browsing features of the present invention.

With respect to the audio interpretation of standard HTML documents, many standard text to speech conversion techniques may be used. The following section describes the techniques which may be used to convert standard HTML documents into audio data. The techniques described herein for converting HTML documents into audio data are exemplary only, and various other techniques for converting HTML documents into audio signals could be readily implemented by one skilled in the art given this disclosure.

Standard text passages are interpreted using conventional text to speech conversion techniques which are well known. The text is interpreted as it is encountered in the document, and such interpretation continues until the user supplies audio input (e.g. to answer a prompt or follow a link), or a prompt is reached in the

document. The end of a sentence is interpreted by adding a pause to the audio, and paragraph marks <p> are interpreted by inserting a longer pause. Text styles may be interpreted as follows.

STYLE	GENERATED AUDIO
<EM>text</EM	Read text with increased volume
<CITE>text</CITE>	Read text as an independent unit (e.g. using inflection and setting off with pauses).
<DFN>word</DFN>	Read text as an independent unit (e.g. using inflection and setting off with pauses).
<CODE>computer code</CODE>	Read punctuation literally and spell out identifiers. If the language of the computer code can be determined, then special reading modes might be applied. For example, C functions might be identified as such.
<KBD>text</KBD>	Read text as usual.
<SAMP>text</SAMP>	Read text as usual.
<STRONG>text</STRONG>	Read text at higher volume.
<VAR>variablename</VAR>	Read variable using a different voice.

Image instructions are specifications in HTML which indicate that a particular image is to be inserted into the document. An example of an HTML image instruction is as follows:

<IMG SRC="http://machine.att.com/image.gif" ALT="[image of car]">

This instruction indicates that the image file "image.gif" is to be retrieved from the machine defined in the URL and displayed by the client browser. Certain conventional graphic browsers do not support image files, and therefore, HTML image instructions sometimes include alternate text to be displayed instead of the image. In the above example, the text "image of car" is included as an alternative to the image file. In accordance with the audio browsing techniques of the present invention, if an image instruction contains a text alternative, then the text is processed and converted to speech and the speech signal is provided to the user. Thus, in this example, the speech signal "*image of* car", would be provided to a user at telephone

110. If no text alternative is provided, then a speech signal is generated indicating that an image with no text alternative was encountered (e.g. "*A picture without an alternative description*").

Conventional HTML contains instructions which support the entering of user input. For example, the following instructions:

```
<SELECT NAME = "selectvar">
<OPTION> mary
<OPTION SELECTED> joe
<OPTION>
</SELECT>
```

request that the user select from two options: mary or joe, with the option joe being selected as a default. In a client executing a conventional graphical browser, these options may be presented, for example, in a pull down menu. In accordance with the audio browsing techniques of the present invention, the above instructions would be translated into speech signals as follows:

*"Please select one of the following: Option mary (pause) Option joe currently selected (pause) end of options. Press *r to repeat these options, press # to continue".*

If the user presses the pound key during the pause after a given option, that option is selected. Whichever item is selected when the user chooses to continue is returned to the document server associated with the variable selectvar. As an alternative to the user making selections with DTMF signals, the user could select the options using voice signals.

Another conventional HTML instruction for entering user input is a checkbox instruction. For example, the sequence of instructions:

```
<INPUT TYPE="checkbox" NAME="varname" VALUE="red" CHECKED>
<INPUT TYPE="checkbox" NAME="varname" VALUE="blue">
<INPUT TYPE="checkbox" NAME="varname" VALUE="green">
```

would result in the following being displayed by a conventional graphics browser:

red ☑
blue ☐
green ☐

The default is that the red box is checked. The user would be able to change this default by checking either the blue or green box. In accordance with the audio browsing techniques of the present invention, the above sequence of instructions would be processed into a speech signal provided to the user as follows:

*The following selections may be toggled by pressing # during the pause: red currently checked (pause), blue (pause), green (pause). Press *r to repeat this list or # to continue.*

By pressing the # key to generate a DTMF signal during a pause, the user can toggle the item preceding the pause. A second press of the # key will move the user out of the input sequence. The user may press *r to repeat the list of options. As an alternative to DTMF audio input, the user may select the checkbox options using voice signal input.

Conventional HTML documents can request user textual input using a TEXTAREA instruction as follows:

<TEXTAREA COLS=60 ROWS=4 NAME="textvar"> Add text here </TEXTAREA>

which, in a conventional graphics browser, would result in the text "Add text here" being displayed followed by a text box of 60 columns by 4 rows being presented to the user for textual input. In accordance with the audio browsing techniques of the

present invention, the above instruction would be interpreted as follows. The COL and ROWS parameters are ignored, and the user is provided with audio:

"Add text here".

The user could then enter DTMF tones followed by the # sign. These DTMF signals would be processed with the results being supplied to the document server associated with the variable "textvar". Alternatively, the user could supply the text by speaking the response into the microphone of telephone 110 and the speech is converted into data by the speech recognition module 214 and the data is supplied to the document server 160 associated with the variable "textvar".

As seen from the above, various techniques can be used such that conventional HTML documents can be browsed in accordance with the audio browsing techniques of the present invention.

In order to more fully exploit the advantages of audio browsing in accordance with the present invention, additional document instructions may be used in addition to the conventional HTML instructions. These instructions, called audio-HTML instructions, may be introduced into conventional HTML documents. These audio-HTML instructions are described below.

A voice source instruction:

<VOICE SRC="//www.abc.com/audio.file">

results in the specified file being played to the user. Such an instruction was described in detail in conjunction with line 512 of the example document 500 of Fig. 5.

A collect name instruction:

<COLLECT NAME="collectvar">

specifies the beginning of a prompt-and-collect sequence. Such a collect name instruction is followed by a prompt instruction and a set of choice instructions. When the user makes a choice, as indicated by audio user input, the results of the user choice are supplied to the documents server associated with the variable collectvar. The collect name instruction, along with an associated prompt-and-collect sequence, is

described in detail in conjunction with the lines 514-524 of the example document 500 of Fig. 5.

A DTMF input instruction:

<INPUT TYPE="DTMF" MAXLENGTH="5" NAME=varname>

indicates that audio user input in the form of DTMF signals is expected from the user. This instruction causes the audio browsing adjunct 150 to pause and wait for DTMF input from the user. The user inputs a DTMF sequence by pressing keys on the keypad of telephone 110 with the end of the sequence indicated by pressing by the # key. The DTMF input is processed as described above in conjunction the example HTML document 500. The decoded DTMF signal is then supplied to the document server associated with the variable varname. The MAXLENGTH parameter indicates the maximum length (DTMF inputs) that are allowed for the input. If the user enters more than the maximum number of DTMF keys (in this example 5), then the system ignores the excess input.

In a similar manner, the SPEECH input instruction:

<INPUT TYPE="SPEECH" MAXLENGTH="5" NAME=varname>

indicates that audio user input in the form of a speech signal is expected from the user. This instruction causes the audio browsing adjunct 150 to pause and to wait for DTMF speech input from the user. The user inputs a speech signal by speaking into the microphone of telephone 110. The speech input is processed as described above in conjunction with the example HTML document 500. The speech signal is then supplied to the document server associated with the variable varname. The MAXLENGTH parameter indicates that the maximum length of the speech input is 5 seconds.

The audio-HTML instructions described herein are exemplary of the types of audio-HTML instructions which may be implemented to exploit the advantages of the audio browsing techniques of the present invention. Additional audio-HTML instructions could be readily implemented by one skilled in the art given this disclosure.

In addition to the above described audio-HTML instructions, the audio browsing adjunct 150 supports various navigation instructions. In conventional graphic browsers, users may use conventional techniques for navigating through a document. Such conventional techniques include text sliders for scrolling through a document, cursor movement, and instructions such as page up, page down, home, and end. In accordance with the audio browsing techniques of the present invention, users may navigate through documents using audio user input, either in the form of DTMF tones or speech, as follows.

DTMF COMMAND	SPEECH COMMAND	NAVIGATION RESPONSE
*8	Top	Jump to beginning of document
*3	End	Jump to end of document
*6	Next	Jump to beginning of next prompt sequence
*7	Skip	Jump to next option, link, definition or other list item
*5	List	List all links within a document with a pause following each link allowing user to specify a selection of the link.

The foregoing Detailed Description is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the Detailed Description, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the principles of the present invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention.

For example, although certain of the communication channels have been described herein as packet switched communication channels, such communications channels could also be implemented as circuit switched communication channels.

What is claimed is:

1. A method for providing audio access to information stored at a server comprising the steps of:

establishing an audio channel between an audio interface device and a telecommunications network node;

establishing a document serving protocol channel between said telecommunications network node and said server;

receiving a document at said telecommunications network node from said server via said document serving protocol channel;

interpreting said received document into audio data at said telecommunications network node; and

transmitting said audio data from said telecommunications network node to said audio interface device via said audio channel.

2. The method of claim 1 wherein said audio interface device is a telephone, said step of establishing an audio channel further comprising the steps of:

receiving a telephone call placed to a telephone number associated with said server;

routing said telephone call to said telecommunications network node.

3. The method of claim 1 wherein said server is a WWW document server and wherein said document serving protocol is hypertext transfer protocol.

4. The method of claim 1 wherein said document includes HTML instructions.

5. The method of claim 4 wherein said document further comprises audio-HTML instructions.

6. The method of claim 1 further comprising the steps of:

receiving at said telecommunications network node audio user input from said audio interface device via said audio channel;

interpreting said audio user input at said telecommunications network node into user data appropriate for transmitting via said document serving protocol; and

transmitting said user data to said server via said document serving protocol channel.

7. The method of claim 6 wherein said audio user input is DTMF tones.

8. The method of claim 6 wherein said audio user input is speech signals.

9. A system for accessing information stored at a server comprising:

a telecommunications network node for receiving a call placed from an audio interface device to a telephone number associated with said server, wherein an audio channel is established between said telecommunications network node and said audio interface device;

a database accessible by said telecommunications network node for associating said telephone number with said server;

means associated with said telecommunications network node for establishing a document serving protocol channel between said telecommunications network node and said server;

an interpreter associated with said telecommunications network node for interpreting a document received from said server via said document serving protocol channel into audio data; and

means associated with said telecommunications network node for transmitting said audio data to said audio interface device via said audio channel.

10. The system of claim 9 wherein said audio interface device is a telephone.

11. The system of claim 9 wherein:

said interpreter is further configured to interpret audio user input received from said audio interface device via said audio channel into user data appropriate for transmission via said document serving protocol; and

said system further comprising means for transmitting said user data to said server via said document serving protocol channel.

12. The system of claim 11 wherein said audio user input is DTMF tones.

13. The system of claim 11 wherein said audio user input is speech signals.

14. The system of claim 9 wherein said server is a WWW document server and wherein said document serving protocol is hypertext transfer protocol.

15. The system of claim 9 wherein said document includes HTML instructions.

16. The system of claim 15 wherein said document further comprises audio-HTML instructions.

17. The system of claim 9 wherein said database comprises data associating telephone numbers with Uniform Resource Locators.

18. A method for providing audio access to information stored at a server which serves documents in accordance with a document serving protocol, said method comprising the steps of:

establishing a communication channel between an audio interface device and said server;

interpreting documents provided by said server into audio data; and

providing said audio data to said audio interface device.

19. The method of claim 18 wherein said step of interpreting takes place at said server.

20. The method of claim 19 wherein said document serving protocol is hypertext transfer protocol.

21. The method of claim 18 wherein said step of interpreting takes place at said audio user interface.

22. The method of claim 18 wherein said step of interpreting takes place at an intermediate node in said communication channel disposed between said server and said audio user interface.

23. The method of claim 18 wherein said document serving protocol is hypertext transfer protocol.

24. The method of claim 18 further comprising the steps of:

interpreting audio user input received from said audio interface device into instructions compatible with said document serving protocol; and

providing said instructions to said server.

25. A system for interpreting information between a server operating in accordance with a document serving protocol and an audio interface device, wherein said server and said audio interface device are connected by a communications channel, said system comprising:

means for receiving a document served by said server via said document serving protocol;

an interpreter for interpreting said received document into audio data; and

means for providing said audio data to said audio interface device.

26. The system of claim 25 wherein said audio interface device is a telephone, said system further comprising means for establishing said communication channel, said means for establishing said communication channel comprising:

means for receiving a telephone call placed from said telephone to a telephone number associated with said server; and

a database for associating said telephone number with said server.

27. The system of claim 25 wherein said interpreter is located at a node disposed between said audio interface device and said server within said communication channel.

28. The system of claim 25 wherein said interpreter is located within said document server.

29. The system of claim 25 wherein said interpreter is located within said audio interface device.

30. The system of claim 25 wherein:

said interpreter is further configured to interpret audio user input received from said audio interface device into instructions appropriate for transmittal in accordance with said document serving protocol; and

said system further comprising means for providing said instructions to said document server.

31. A document server for providing audio access to stored documents comprising:

an interface for connection with a communication link, said communication link providing communication with an audio interface device;

a machine readable storage device storing computer program instructions and said documents;

a central processing unit connected to said memory and said interface for executing said computer program instructions, said computer program instructions causing the central processing unit to perform the steps of:

in response to receipt of a request for a document, retrieving said requested document from said machine readable storage device in accordance with a document serving protocol;

interpreting said requested document into audio data; and

transmitting said audio data to said audio interface device via said interface.

32. The document server of claim 31 wherein said document serving protocol is hypertext transfer protocol

33. The document server of claim 31 wherein said communication link is a telephone network connection, said document server further comprising:

a telephone network interface.

34. The document server of claim 31 wherein said communication link is a packet network connection, said document server further comprising:

a packet network interface.

35. The document server of claim 31 wherein said computer program instructions further cause the central processing unit to perform the steps of:

in response to audio user input received from said audio interface device via said communication link, interpreting said audio user input into user data; and

in response to said user data, retrieving a document from said machine readable storage device in accordance with a document serving protocol.

FIG. 1

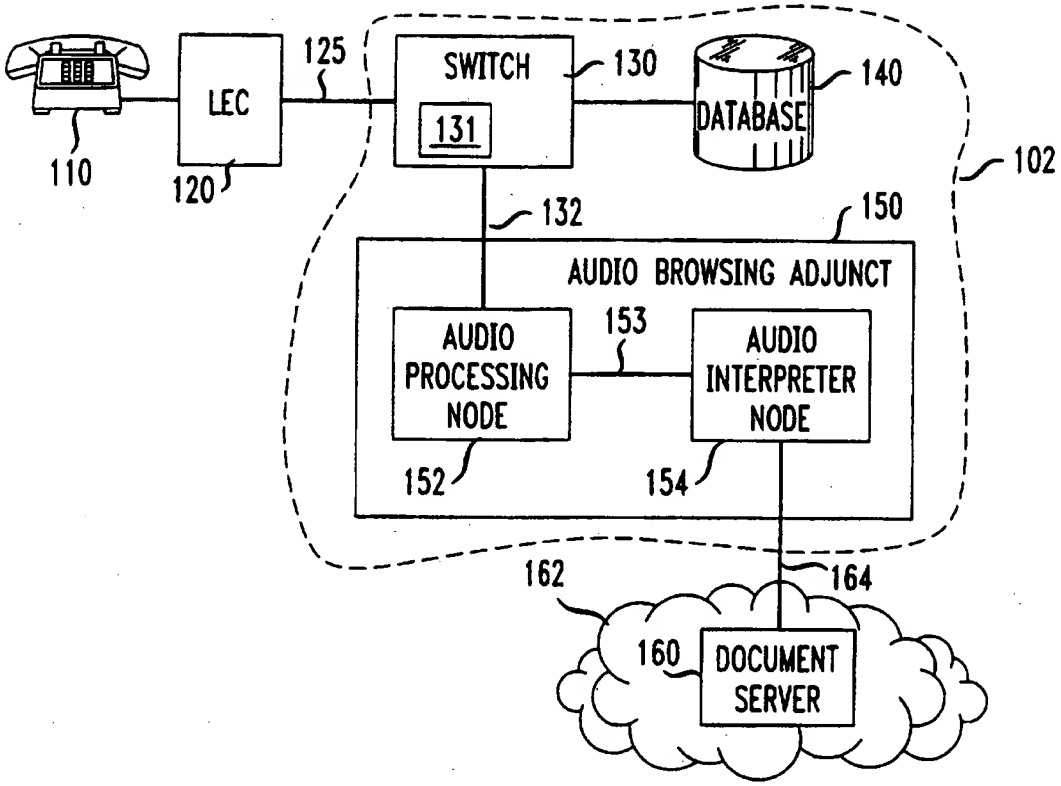

100
1/8
LEC
125
SWITCH
130
131
DATABASE
140
110
120
102
132
150
AUDIO BROWSING ADJUNCT
AUDIO PROCESSING NODE
153
AUDIO INTERPRETER NODE
152
154
162
164
160
DOCUMENT SERVER

FIG. 2

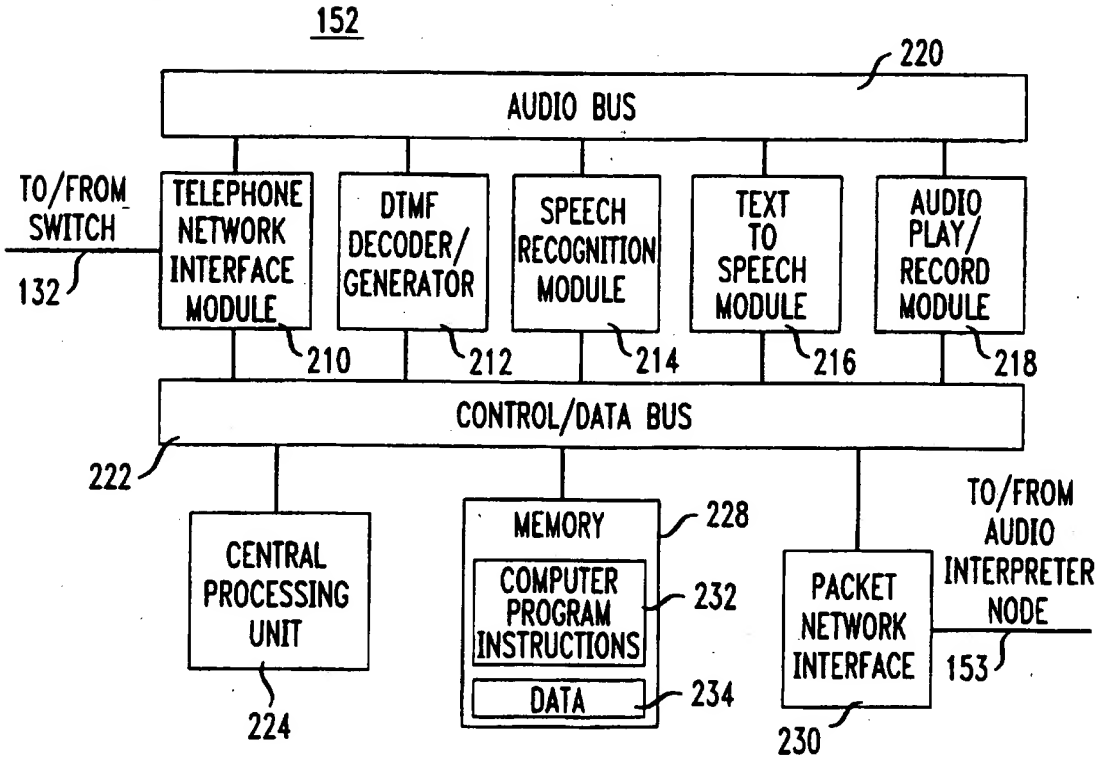

152
220
AUDIO BUS
TO/FROM SWITCH
132
TELEPHONE NETWORK INTERFACE MODULE
DTMF DECODER/ GENERATOR
SPEECH RECOGNITION MODULE
TEXT TO SPEECH MODULE
AUDIO PLAY/ RECORD MODULE
210
212
214
216
218
CONTROL/DATA BUS
222
CENTRAL PROCESSING UNIT
224
MEMORY
228
COMPUTER PROGRAM INSTRUCTIONS
232
DATA
234
PACKET NETWORK INTERFACE
230
TO/FROM AUDIO INTERPRETER NODE
153

FIG. 3

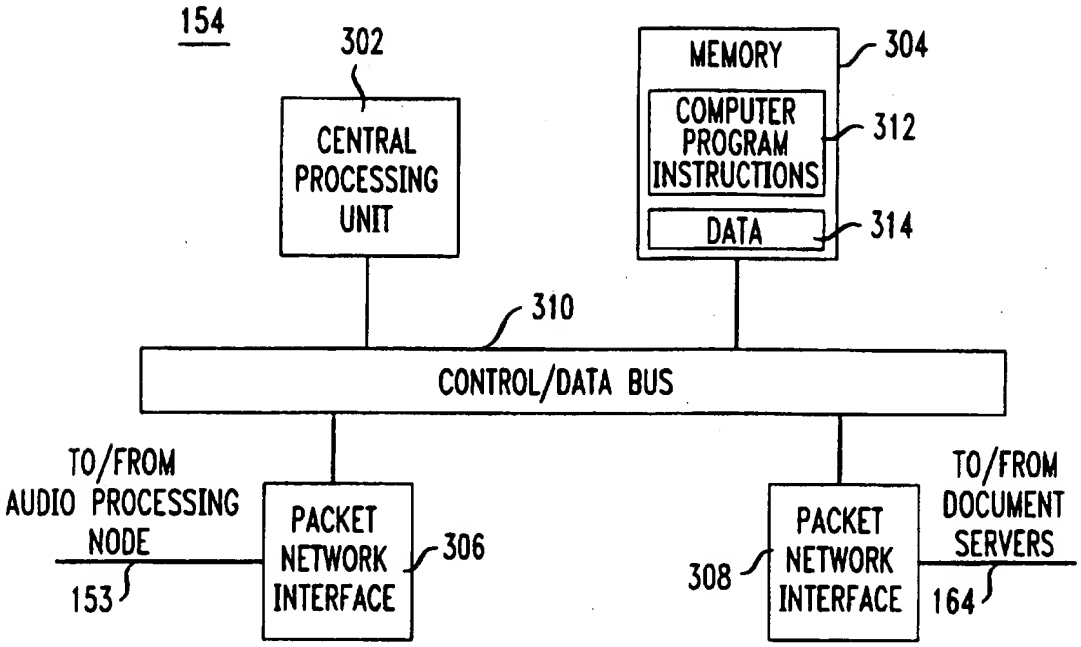

154
302
CENTRAL PROCESSING UNIT
MEMORY
304
COMPUTER PROGRAM INSTRUCTIONS
312
DATA
314
310
CONTROL/DATA BUS
TO/FROM AUDIO PROCESSING NODE
153
PACKET NETWORK INTERFACE
306
308
PACKET NETWORK INTERFACE
TO/FROM DOCUMENT SERVERS
164

FIG. 4

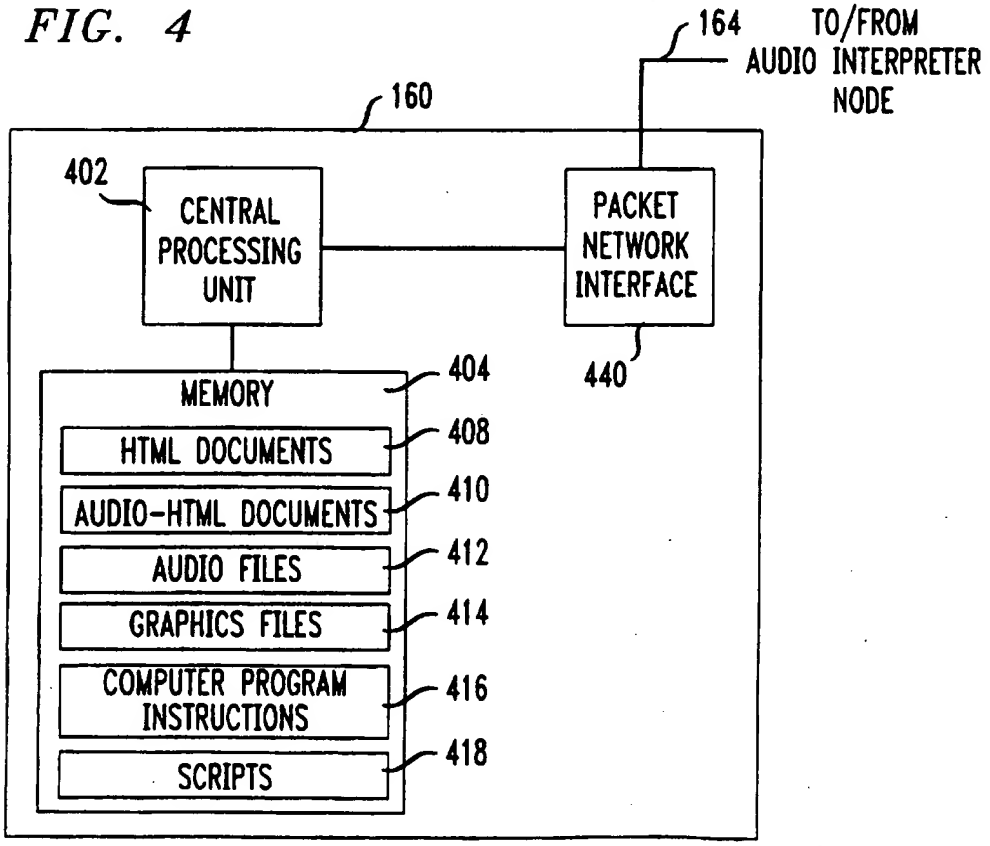

164
TO/FROM AUDIO INTERPRETER NODE
160
402
CENTRAL PROCESSING UNIT
PACKET NETWORK INTERFACE
440
MEMORY
404
HTML DOCUMENTS
408
AUDIO-HTML DOCUMENTS
410
AUDIO FILES
412
GRAPHICS FILES
414
COMPUTER PROGRAM INSTRUCTIONS
416
SCRIPTS
418

FIG. 5

500

```
    <HTML>
502 <HEAD>
504 <TITLE>Greeting</TITLE>
506 </HEAD>
    <BODY>
508  Hello!
510<FORM METHOD=GET ACTION="http://machine:8888/hastings-bin/getscript.sh">
512 <VOICE SRC="//www-spr.ih.att.com/~hastings/annc/greeting.mu8">
514 <COLLECT NAME="collectvar">
516<PROMPT><VOICE SRC="http://www-spr.ih.att.com/~hastings/annc/choices.mu8">
      <PROMPT>
518 <CHOICE VALUE="Joe" SEQUENCE="1">
520 <CHOICE VALUE="Jim" SEQUENCE="2">
522 <CHOICE VALUE="Bob" SEQUENCE="3">
524 </COLLECT>
   </FORM>
   </BODY>
   </HTML>
```

FIG. 6

600

```
    <HTML>
    <HEAD>
    <TITLE> Page of links</TITLE>
    </HEAD>
    <BODY>
602 This page gives you a choice of links to follow to other world wide web pages. Please
603 click on one of the links below.
604   <A HREF="http://www.abc.com/cars.html">click here for information on cars </A>
605   <A HREF="http://www.abc.com/trucks.html"/>click here for information on trucks
      </A>
610 You may also skip an advertisement and get a health tip by following this link.
620 <A HREF="#endofpage">here.</A>
    This message brought to you by your favorite company.
625 <A NAME="endofpage">One final word:</A>
    An apple a day keeps the doctor away.
    </BODY>
    </HTML>
```

FIG. 7

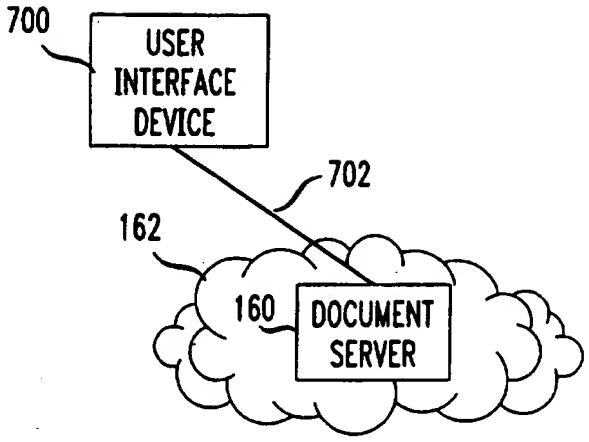
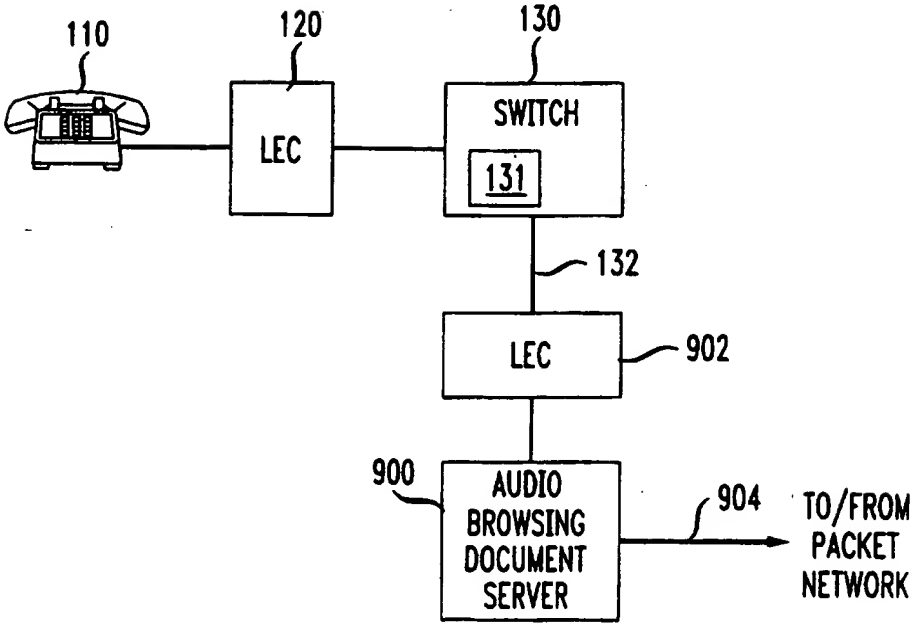

700
USER INTERFACE DEVICE
702
162
160
DOCUMENT SERVER

FIG. 9


110
120
130
LEC
SWITCH
131
132
LEC
902
900
AUDIO BROWSING DOCUMENT SERVER
904
TO/FROM PACKET NETWORK

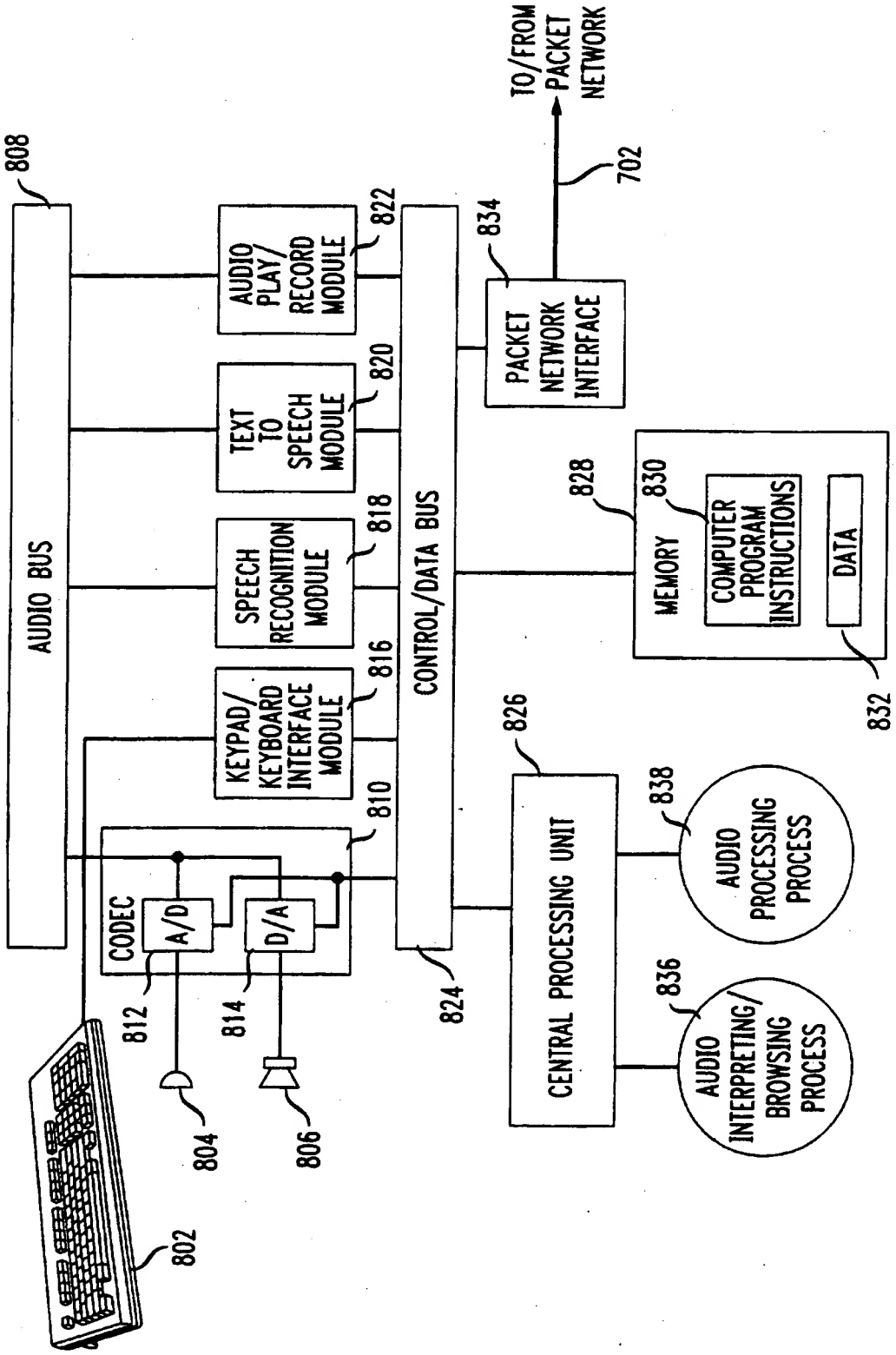

700
AUDIO BUS
808
AUDIO PLAY/ RECORD MODULE
822
TEXT TO SPEECH MODULE
820
SPEECH RECOGNITION MODULE
818
KEYPAD/ KEYBOARD INTERFACE MODULE
816
CODEC
A/D
D/A
810
812
814
804
806
802
CONTROL/DATA BUS
824
PACKET NETWORK INTERFACE
834
TO/FROM PACKET NETWORK
702
MEMORY
828
830
COMPUTER PROGRAM INSTRUCTIONS
DATA
832
CENTRAL PROCESSING UNIT
826
AUDIO PROCESSING PROCESS
838
AUDIO INTERPRETING/ BROWSING PROCESS
836

FIG. 8

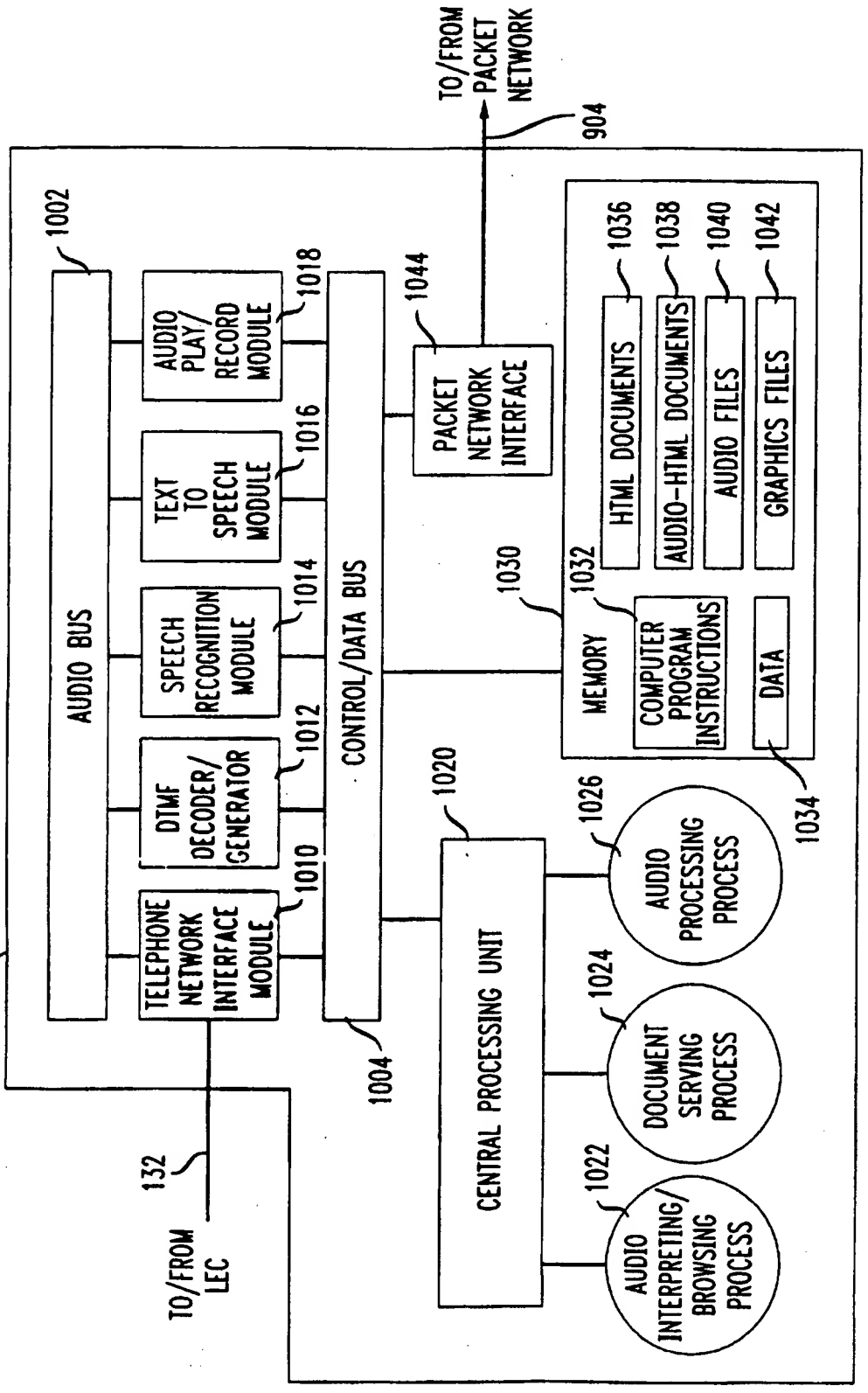

900
TO/FROM LEC
132
TELEPHONE NETWORK INTERFACE MODULE
1010
DTMF DECODER/ GENERATOR
1012
SPEECH RECOGNITION MODULE
1014
TEXT TO SPEECH MODULE
1016
AUDIO PLAY/ RECORD MODULE
1018
AUDIO BUS
1002
CONTROL/DATA BUS
1004
CENTRAL PROCESSING UNIT
1020
AUDIO INTERPRETING/ BROWSING PROCESS
1022
DOCUMENT SERVING PROCESS
1024
AUDIO PROCESSING PROCESS
1026
PACKET NETWORK INTERFACE
1044
TO/FROM PACKET NETWORK
904
MEMORY
1030
1032
COMPUTER PROGRAM INSTRUCTIONS
DATA
1034
HTML DOCUMENTS
1036
AUDIO-HTML DOCUMENTS
1038
AUDIO FILES
1040
GRAPHICS FILES
1042

FIG. 10

FIG. 11

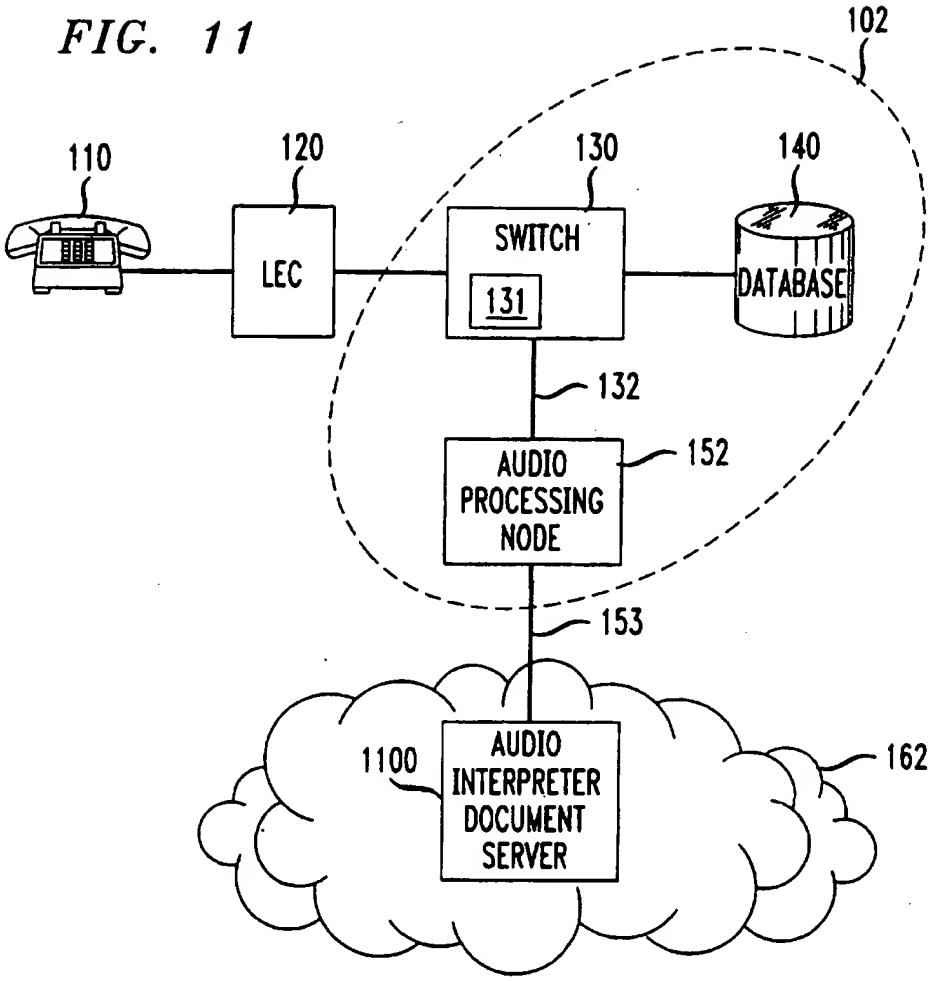
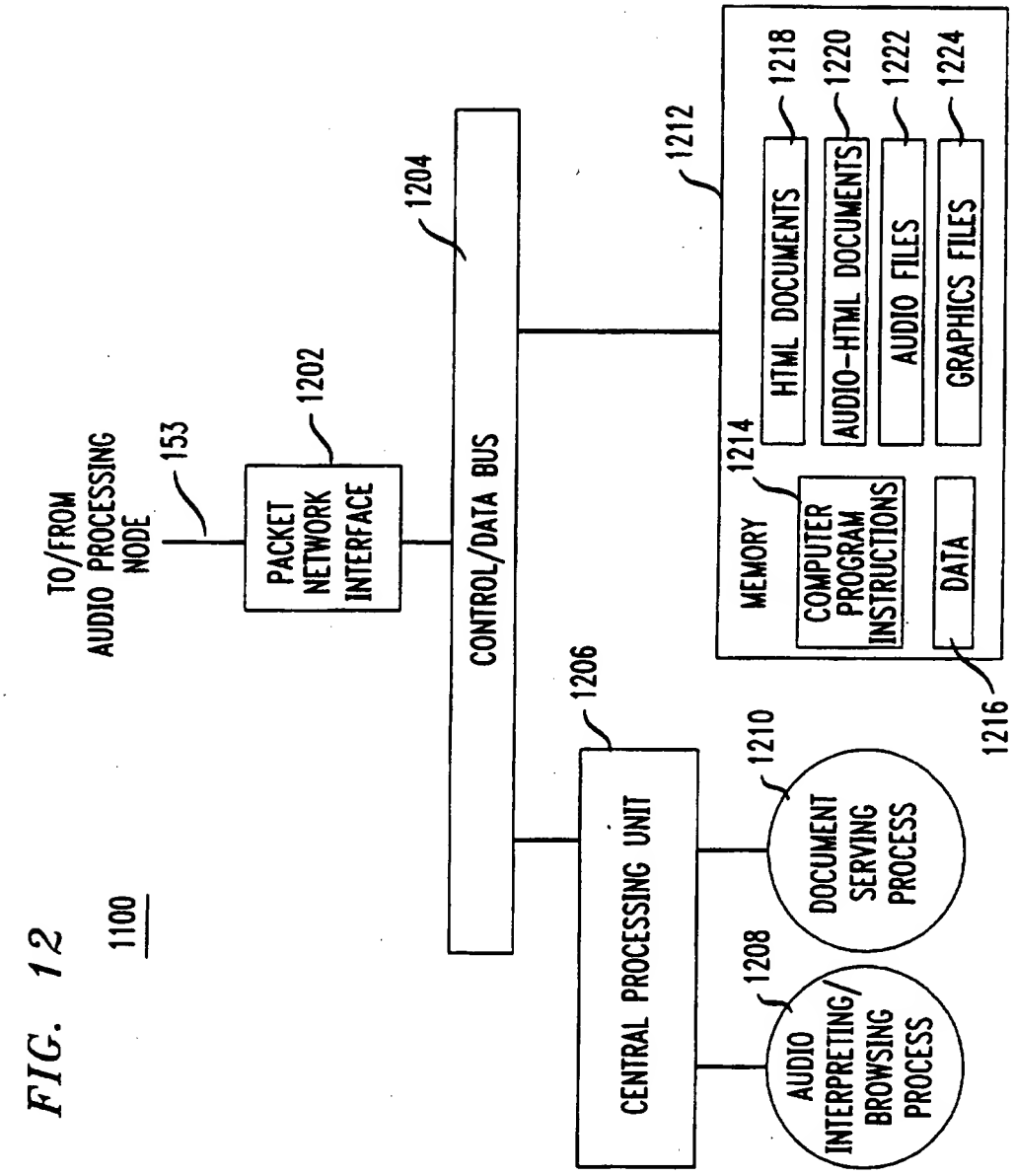

102
110
120
130
140
LEC
SWITCH
131
DATABASE
132
AUDIO PROCESSING NODE
152
153
1100
AUDIO INTERPRETER DOCUMENT SERVER
162

FIG. 12


1100
TO/FROM AUDIO PROCESSING NODE
153
PACKET NETWORK INTERFACE
1202
CONTROL/DATA BUS
1204
CENTRAL PROCESSING UNIT
1206
AUDIO INTERPRETING/ BROWSING PROCESS
1208
DOCUMENT SERVING PROCESS
1210
MEMORY
1212
1214
HTML DOCUMENTS
1218
AUDIO-HTML DOCUMENTS
1220
AUDIO FILES
1222
GRAPHICS FILES
1224
COMPUTER PROGRAM INSTRUCTIONS
DATA
1216

INTERNATIONAL SEARCH REPORT

International ication No
PCT/US 97/03690

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04L29/06 H04M3/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04L H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BT TECHNOLOGY JOURNAL, vol. 14, no. 1, 1 January 1996, pages 57-67, XP000554639 PAGE J H ET AL: "THE LAUREATE TEXT-TO-SPEECH SYSTEM - ARCHITECTURE AND APPLICATIONS"	1,6,9, 11,18, 22,24, 25,27, 31,33-35
Y	see abstract	2-4,7,8, 10, 12-16, 26,30,32
A	see paragraph 2.2 - paragraph 3.2	5,17, 19-21, 23,28,29

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
18 September 1997	25.09.1997

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cichra, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International A :ation No
PCT/US 97/03690

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY (EUROSPEECH), PARIS, SEPT. 26 - 28, 1989, vol. 1, 26 September 1989, TUBACH J P;MARIANI J J, pages 561-564, XP000209922 RICCIO A ET AL: "VOICE BASED REMOTE DATA BASE ACCESS"	1,2,6, 9-11,18, 19, 24-26, 28,30, 31,33-35
Y	see abstract	23
A	see paragraph 1. - paragraph 3.1	3,4,8, 13,14, 17,20,32
	see paragraph 3.3 see paragraph 4.2 see paragraph 5.2 ---	
X	CSELT TECHNICAL REPORT ON EUROSPEECH 1991. MARCH 1992 REPORT CONTAINS C.D. AT BACK OF ISSUE, vol. 20 - NO 1, 1 March 1992, TOSCO F, pages 79-83, XP000314315 BAGGIA P ET AL: "A MAN-MACHINE DIALOGUE SYSTEM FOR SPEECH ACCESS TO E-MAIL INFORMATION USING THE TELEPHONE: IMPLEMENTATION AND FIRST RESULTS"	1,2,6, 9-11,18, 19, 24-27, 30,31, 33-35
A	see abstract	3,4,8, 13,14, 17,20, 23,32
	see paragraph 1.1 - paragraph 2.2 see paragraph 2.6 ---	
Y	COMPUTER NETWORKS AND ISDN SYSTEMS, 1 December 1995, pages 53-59, XP002037371 GESSLER, KOTULA: "PDAs as mobile WWW Browsers"	3,4,7,8, 12-15, 23,30,32
A	see the whole document	1,2,5,6, 9-11, 16-22, 24-26, 31,33-35

P,X	BELL LABS TECHNICAL JOURNAL, vol. 2, no. 1, 1 January 1997, pages 19-35, XP002036350 ATKINS D L ET AL: "INTEGRATED WEB AND TELEPHONE SERVICE CREATION"	1-15,18, 22, 24-27, 30-35
A	see page 19, line 1 - page 21, last line	17,19, 21,23
P,Y	see page 28, column 1, line 11 - page 30, column 1, line 23 see page 33, column 1, line 3-15 see page 34, column 1, line 9-20 ---	16
	-/--	

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International / cation No
PCT/US 97/03690

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	AT & T TECHNICAL JOURNAL, vol. 69, no. 5, 1 September 1990, pages 61-76, XP000224080 FISCHELL D R ET AL: "INTERACTIVE VOICE TECHNOLOGY APPLICATIONS"	2,10,26
A	see the whole document	1,6,7,9, 11,12, 18,19, 24,25, 27,30, 31,33-35

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)